= grammar/spelling

= conceptual/understanding

Introduction

This paper discuses Donald Davidson's critique ([1990b] 2004) of the Turing Test as an illustration of Davidson's theory of linguistic and intellectual competence. After a short introduction to Donald Davidson it consists of three main sections.

Section 1 explores Turing's original ideas and their relevance. It discusses different interpretations and versions of the Turing Test and establishes wich interpretation I follow. It also contains a discussion of the significance of the test and Turing's ideas about Machine Learning for the field of Artificial (General) Intelligence.

Section 2 presents and evaluates where Davidson agrees and disagrees with Turing. It also presents Davidson's modified version of the Turing Test which allows to judge the computer based on an observation of its learning process. I argue that this test acts as a definition for intelligence and expression of Davidsons theory of linguistic and intellectual competence.

Section 3 delves deeper into Davidson's theory by looking at its parts: reconstructing his argument for supervenience of the mental, exposing his epistemology of triangular knowledge, and describing his empirical anti-conventionalist theory of interpretation. It shows how those elements come together as a unified theory about the essence of linguistic and intellectual abilities that constitute Davidson's definition of rationality or intelligence that is incorporated in Davidson's Test. After that I discuss how Davidson's Test fences against some classical objections against the Turing Test and some other concerns.

I conclude with a summary of the core agreements and distinctions between Davidson and Turing and a brief discussion of the implications of Davidson's ideas (holistic theory and anti-conventionalism) for Computer Linguistic and Machine Learning approaches that might deserve further investigation.

Introduction to Donald Davidson

Donald Davidson (1917–2003) "was one of the most important philosophers of [action, language, and mind in] the latter half of the twentieth century [...] with a reception and influence that, of American philosophers, is perhaps matched only by that of [his friend, mentor, and colleague] W. V. O. Quine" (Malpas [1996] 2015, p. 1). His thoughts are exposed through a collection of very densely written essays (published from 1963–2001)[1] which "form a mosaic out of which emerges a unified and surprisingly elegant overall view of the mind and its relation to the world. It sees our *nature as linguistic beings* as the key to the possibility of thought, to the objectivity of the world [...], and to how the [irreducably mental] mind moves us to action in a [physical] world" (Lepore and Ludwig 2009, p. 1, emphasis added)

To give a very brief overview of Davidson's philosophy:

Figure 1 – Donald Davidson by David Levine in Forum Gallery, New York

- He takes the causal nature of thought ([1963] 2001) as an argument for the supervenience and irreducibility of the mental to the physical known as 'Anomalous Monism' ([1970] 2001).
- He argues that semantic theories for natural languages must be recursive ([1965] 2001) and that meaning and propositions can be replace by truth conditions and distal stimuli ([1967] 2001).
- Based on Quine's Radical Translation, he takes interpretation as a process of adapting a theory of intertwined understanding and believe about the speaker ([1973b] 2001).
- He rejects Conventionalism ([1984] 2001) based on his argument for the primacy of ideolect, placing the social aspect of language in Epystemology ([1994] 2005).
- Rejecting positivist and cartesian Epistemology, he sees knowledge of self, world, and others as interdependent ([1991b] 2001).
- His Unified Theory ([1980] 2004) aims to captures the essence ([1995a] 2004) of mind, language, and action.
- He compares his philosophy to Spinoza ([1999] 2005) and Plato and Gadamer ([1997] 2005).

HISTORICAL CONTEXT

Language has been a philosophical topic since Plato, but it especially gained attention in the so-called 'Linguistic Turn' of the 20th century, when it was recognized as the medium for our rational access to the world. Names like Frege, Russel, the early Wittgenstein, and the Vienna Circle stand for the earlier Ideal Language Philosophy and Logical Empiricism which hoped to solve problems of Metaphysics through a *definite logical analysis* of meaning. However, This approach did not turn out very successful. Led by Austin, Ryle, and the late Wittgenstein, analytical philosophy paid more attention to Ordinary Language, focusing on the importance of *social and performative aspects*. Quine and Davidson stand at the resolution of this tension with an approach that recognizes linguistic behavior in its context while leveraging the formal methods for their analysis (Bertram 2011). Through Davidson the analytical tradition also became linkable to the 'continental' Hermeneutic traditions, contributing to the dissolution of the philosophical divide of the 20th century. (Malpas [1996] 2015, p. 2)

1 – It is remarkable that a contemporary philosopher of Davidson's calibre only started publishing his relevant papers in his 40s after he had been a professor for over a decade. The publication dates of Davidson's papers ought to be interpreted with care though. His fried and student Ernie Lepore recounts: "[Davidson was] adventurous and daring [...] from early on right up until his death. [...] Donald was without even the slightest speck of careerism from the very start. He traveled the world giving papers in exotic places and often handed them over to local journals upon request. [...M]any of the papers were written somewhat contemporaneously. [...Some were] given much earlier than they were published and relatively around the same time. [...] I possess a mimeographed copy of his quotation paper dated from the early 60's, long before its 1979 publication." [Lepore:2003me].

Davidson's life and career are actually quite inspiring, *fascinating*, and unconventional. Davidson started out in the History of Philosophy and ventured into psychology and economy during his studies and as a professor. Certainly it is no coincidence that he was cited by continental and analytical philosophers and even computer scientists alike. To learn more about Davidson's life I recommend to read Lepore's 'Interview with Donald Davidson' ([1999] 2004) and Davidson's own 'Intellectual Autobiography' [Davidson:1999ia]

1 The Turing Test and Its Relevance

Alan Turing (1912–1954) first proposed his famous test in 'Computing Machinery and Intelligence' (1950). The thesis of his paper is that the question "Can machines think?" — which he deemed too vague to deserve discussion — can be replaced with the question wether a computer can pass a specified test (that is now known as the 'Turing Test'). This test is described as a game: the imitation game.

The imitation game consists of an *interrogator* who can communicate via a real-time text chat interface with two players. One player is a *computer*, the other a *human*. The interrogator's task is to identify the human after a given time. The computer's task is to pretend to be a human[2] and trick the interrogator into wrongly identifying it as the human. The human's task is to help the interrogator to make the correct identification. If many interrogators consistently[3] can't tell the computer apart from the human the computer wins the game and passes the test. (Turing 1950, p. 433–434)

In the following I discuss different interpretations of the Turing Test and establish which interpretation and version of the test I take as the basis for my discussion. Furthermore, I argue for the relevance of this interpretation of the test and work out 4 main claims from Turing's discussion of his test. In doing so I specifically emphasize some of Turing's visionary ideas on how a computer must be programmed to pass the test, that have implications for Davidson's discussion of the test.

1.1 Interpretations and Versions of the Turing Test

The interpretations are not as straight forward as it might seem. There are three main interpretations for what it means to pass the Turing Test:

1. The Turing Test is taken as an operational *definition* — a *sufficient* and *necessary* condition — of *intelligence*[4]: Something is intelligent if and only if it passes the Turing Test. (Such an interpretation is for example found in Millar 1973.)
2. The Turing Test is taken as a *sufficient* condition of intelligence: Something is intelligent if it passes the Turing Test. But it is not necessary for something to pass the Turing test in order to be intelligent. (Davidson [1990b] 2004 takes this interpretation.)
3. The Turing Test is taken "as a potential source of good inductive evidence for" intelligence: If something passed the Turing Test one would be *justified* for inductively inferring that it is intelligent. (This interpretation goes back to Moor 1976, p. 249, 251.)

The first interpretation and any reading of intelligence in a broader sense are clearly not in line with Turing's ideas. Turing was not interested in a definition of intelligence but in setting a clear goal for further research. In fact I would argue that the third interpretation from Moor does Turing most justice. This becomes fairly clear from the following section of a radio interview ([1952] 1999), which represents Turing's simplest expression of the test:

> "I *don't want to give a definition of thinking* [...] I don't really see that we need to agree on a definition at all. The important thing is to try to *draw a line between the properties of a brain, or of a man, that we want to discuss, and those that we don't.* [...] I would like to suggest a particular kind of test that one might apply to a machine. You might call it a test to see whether the machine thinks, but it would be better to avoid begging the question, and *say that the machines that pass are (let's say) 'Grade A' machines.* The idea of the test is that the machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretence is reasonably convincing." (Turing [1952] 1999, p. 466, emphasis added)

Turing clearly describes the imitation game as a test that defines a special *class of computers*, but not as one that defines thinking/intelligence. It is also implicit that he sees the *ability to communicate* as quintessential for human-like intellectual abilities.

The quotation also gives a simplified version of the test, in so far as it removes the second human player. While this might be limiting for a quantitative analysis, it ensures the focus on the main question and removes any undue emphasis on strategy. Davidson proposes the same simplification in his discussion ([1990b] 2004). He goes even further in proposing that since "we might count an object as thinking even if it were easily distinguishable from a person. [...T]he interrogator [...should simply] be asked to decide whether or not the object is thinking" ([1990b] 2004, p. 81). If we assume that any bias against the object's ability to think can be removed (which makes sense for this philosophical discussion), this seems to be the clearest expression of Turing's idea. Therefore, I will follow Davidson and refer to this simplified version in the following.

1.2 Relevance of the Test and the Turing Principle

Where does it leave the relevance of the Turing Test, if we follow Moor and interpret it merely as a framework to gather evidence for human-like intelligence? For Computer Scientists it might be *prudent* to follow Turing and ignore the philosophical question of a definition of intelligence and instead take the Turing Test as a definition of a certain class of computers. In this way the Turing Test can give a clear goal to the field of 'Artificial General Intelligence' (AGI), for which the unclarity about a clear definition of its topic is a core challenge. Unlike other propositions for an operational definition[5] of AGI, the Turing Tests provides a clear and well-justified empirical goal. In his analysis Moor argues why the Turing Test is well apt for that:

> "[T]here are two strong arguments why the Turing test is a good format for gathering inductive evidence. First, the Turing test

2 – I think it is pretty clear that Davidson ([1990b] 2004, p. 78) misinterprets Turing when he suggests that Turing wants the computer to play the literal imitation game and pretend to be a *woman* and that the interrogator's task would be to decide on the gender. This becomes pretty clear when Turing says: "If the man were to try and pretend to be the machine [...]" (1950, p. 435) and from all his examples that are focused on how the machine can imitate being a *human* not a woman. (See also Copeland 2000, p. 526.)

3 – Turing's idea is that a baseline is established by the traditional imitation game, in which a man tries to imitate a woman and the interrogator has to decide on the gender. While this might not be the best way to establish a baseline, it means that Davidson's critique that "Turing does not say what he would make of a computer that was consistently chosen over the [...human] to be the [...human]" ([1990b] 2004, p. 78), is ill-conceived, as the computer will unambiguously fall over or under the baseline within a margin of error.

4 – The term 'intelligence' is adopted here purely in reference to Artificial Intelligence. Turing mostly uses the term 'thinking' and Davidson ([1995a] 2004) mostly uses the term 'rationality' to refer to the concept that this thesis is concerned with. I will introduce and explicate the term intellectual competence for this concept in subsection 3.4.

5 – See Muehlhauser (2013) for a brief overview of operational definitions of AGI, including the Coffe-Brewing and College Test. From a philosophical perspective all those seem rather random and dubious — certainly much further away from clearly capturing necessary conditions of human-like intellectual competences.

permits direct or indirect testing of virtually all of the activities one would count as evidence for thinking. Secondly, the Turing test encourages severe testing. [...T]he computer would be tested in detail over a wide range of subjects [...and] the interrogator's goal is to find a refuting instance which gives the computer away." (Moor 1976, p. 251–252)

Critics of the Turing Test as a goal for AGI mostly fall into two camps. The first point Moor provides is aimed against the first type of critic who suggests the test sets the wrong goal. As Moor argues, communication is a very clear framework to investigate all kinds of thinking. A further practical reason for the relevance of the Turing Test — that I would add — is that natural language communication provides a 'gold standard' for completely natural and seamless computer user interfaces. Many of those critics mistake the question as philosophical while it is best treated as definitional. Computer Scientists ought not to be concerned with defining intelligence in general but with a good definition for "Grade A" computers — to use Turing's terminology. In this regard, the interpretation of the Turing Test as a framework to gather empirical evidence can be set as a clear definition of the goal of AGI.

The other type of critic questions the adequacy of the test to determine whether the goal is reached. Moor's second point is aimed against that by pointing out how well the test encourages thorough testing. The critics sometimes mistake Turing's predictions as a specification that the test ought to take only 5 minutes and limited implementations of the Turing Test that favor engineering tricks like the Loebner Prize[6] discredit Turing's intentions. I agree with Copeland's interpretation of Turing: the goal set by the test is a computer that "plays the imitation game successfully come what may, with no field of human endeavour barred, and for any length of time commensurate with the human lifespan." (Copeland 2000, p. 530)

TURING'S THREE MAIN CLAIMS

Following Copeland (2000, p. 530), we can establish the following two main claims of Turing from our previous discussion:

1. Turing sees communication as suitable to expose the relevant intellectual abilities that determine whether a computer can perform tasks on par with a human being.
2. Turing thinks that his test *specifically* can determine whether a computer possesses such communicative abilities.

But there is also a third claim which is known as the Turing Principle[7]:

3. Turing believes that universal computers can simulate any physical process, including the brain.

This might be most clearly expressed in Turing's lecture on 'Can Digital Computers Think?': "If it is accepted that real brains [...] are a sort of machine it will follow that our digital computer suitably programmed will behave like a brain" (Turing [1951] 1999, p. 463). This is a much more controversial claim and — as Turing was well aware — he had few arguments and even less evidence for it.

It is important to highlight that this claim is not central in Turing's writing and that it is separate and fully *independent* from the others claims. Nevertheless, it is the main point many philosophers have attacked. (The most famous example is probably Searle 1980.) This is often tied to the misinterpretation of the Turing Test as a definition or sufficient condition for intelligence in a philosophical sense. I am not particularly interested in the discussion of this third claim here.

1.3 Turing on Machine Learning

Unlike the obsession in popular culture with human-like 'Artificial Intelligence' (AI) might suggest, the interest of Computer Scientists in the last years has been more focused on so-called 'weak' AI. This refers to applying 'Machine Learning' to apply domain-specific intelligence to special problems (most notably: image recognition, domain-specific language processing, and robotics). The field that is concerned which this is referred to as 'Artificial Intelligence' these days. The field that is concerned with human-like strong AI is the field of Artificial General Intelligence (AGI) which was referenced in the previous section. This shift of focus has mainly been due to a lack of success in AGI.

Turing believed that by the turn of the century there would be computers that "play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning" (1950, p. 442). Recent winners of the Loebner Prize (AISB 2015), which is awarded to the most human-like chatbot judged by a Turing-inspired test, show that this has not quite come to pass. On the other hand, Turing himself said that it would take "*at least* 100 years" ([1952] 1999, p. 434) until a general Turing Test could be passed — and the goal of AGI reached — which is still well within the realm of the possible. Turing's wrong prediction of the development can be mainly accredited to his underestimation of the required processing power for Machine Learning (1950, p. 455). Recent accomplishments in AI with deep neural networks, for example, have only become feasible because computers are able to run networks with millions of neurons in real time.[8]

TURING'S FOURTH MAIN CLAIM

However, Turing's idea about *how* to build a computer that could pass his test are more interesting than his timeline predictions. His claim that "the problem [of building a computer that passes the Turing Test] is mainly one of programming" (1950, p. 455) still rings true today.

We have already learned of his idea that computers should be able to simulate the brain ([1951] 1999). Indeed this reverse engineering approach to AI is the basic idea behind the neural networks

6 – Shieber (1994) criticizes the Loebner Prize for its inappropriateness to award advances in natural-language-processing techniques instead of engineering tricks oriented to the exigencies of the restricted task like parrying and insertion of random typing errors. The setup of the scoring system alone shows how pointless it is to even judge current systems by a direct Turing Test. He argues that a subjective award modeled after the Nobel Prizes would make significantly more sense.

7 – This is also known as the 'Church–Turing–Deutsch Principle' and represents an extension of the well-known Church-Turing Thesis to artificial intelligence.

8 – Steven Wolfram writes: "Computers (and especially linear algebra in GPUs) got fast enough that [...] it became practical to train neural networks with millions of neurons, on millions of examples. [...T]his suddenly brought large-scale practical applications within reach. [...] I don't think it's a coincidence that this happened right when the number of artificial neurons being used came within striking distance of the number of neurons in relevant parts of our brains. [...I]f we're trying to achieve 'human-like' image identification [...] then this defines a certain scale of problem, which, it appears, can be solved with a 'human-scale' neural network." (2015)

which power the most successful image entity-recognition algorithms today. Turing ([1948] 1992) pioneered this approach with his B-type unorganized machine which consisted of neurons that are trained through an 'education' process, and he proved that they were equivalent with digital computers (see also Copeland and Proudfoot 1999). It is plausible that we might achieve AGI through brain simulation before we even deeply understand how the brain works.

Even more interesting is that Turing also predicted the Machine Learning approach to build intelligent algorithms and saw its non-deterministic nature as a characteristic feature:

> "Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. [...] We have thus divided our problem into two parts. The child-programme and the education process. These two remain very closely connected. We cannot expect to find a good child-machine at the first attempt. One must experiment with teaching one such machine and see how well it learns." (Turing 1950, p. 456)

> "An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behaviour. [...] This is in clear contrast with normal procedure when using a machine to do computations : one's object is then to have a clear mental picture of the state of the machine at each moment in the computation. [...] Processes that are learnt do not produce a hundred per cent certainty of result; if they did they could not be unlearnt." (Turing 1950, p. 458–459)

This description is very much how modern statistical Machine Learning algorithms work. They consist of an algorithm that describes a mathematical model which is trained with human-annotated data (for example, a grammar analysis of sentences) and are then able to perform their task on similar data. (See Schubert [2014] 2015 for an overview about approaches to Natural Language Processing; and Jurafsky and Martin 2015 for a more technical introduction.) As Turing mentions, this is a paradigm shift from classical algorithms which have results that are clearly defined by the programmer and predictable independent of any training data. We might add a fourth core claim:

4. Turing believes that the approach to device intelligent computers should be based on learning algorithms that are trained and do not behave ~~predictable~~ in a classic sense.

Note that devising intelligent algorithms is a quite different task and involves techniques where an interpretation of the state of the program at each step becomes difficult or even impossible — especially in the case of deep neural networks. And where the outcome is not just dependent on the set of rules specified in the programming but also on the 'experience' gathered in the program during its learning process. We will see later why this matters for a Davidsonian perspective.

SUMMARY

We have seen that a simplified interpretation of the Turing Test as proposed by Davidson ([1990b] 2004) reveals the core of Turing's idea. However, we have also seen that an interpretation of the Turing Test as an operational definition or sufficient condition for intelligence is not in line with Turing's writing. Instead, I have proposed to follow Moor (1976) and interpret the test as a framework to collect empirical evidence to show that a computer can perform human-like tasks. We have seen that this can provide a good and clear definition of a special class of computers that are the goal of AGI research.

Furthermore, we have established that Turing claims that (1) communication abilities are representative of intellectual abilities in general and (2) that his test is adequate to evaluate those abilities. I have pointed out that his claim (3) that computers can simulate the brain (the Turing Principle), which is the main point most philosophers critique, is completely independent of the other claims and is the least essential one for Turing.

Lastly, we have seen that Turing had pioneering ideas about how to devise algorithms which could pass his test. Such algorithms (4) need to be able to learn when trained with data and, different from classical algorithms, their outcome is not predictable independent of their 'experience' and the states of their operation are not easily interpretable.

2 Davidson's Critique of the Turing Test

In this section I will first discuss the relevance of the Turing Test for Philosophy of Mind and for Davidson specifically. I will then argue where Davidson agrees with Turing, where his criticism of Turing applies, and in how far I think this criticism does Turing justice. The section ends in an exposition of Davidson's proposal for a modified Turing Test that provides a definition for intelligence.

2.1 Relevance of Turing's Test for Davidson's Philosophy of Mind

I have argued that the Turing Test is misunderstood as a definition or sufficient condition for intelligence in general and merely provides a pragmatic goal for AGI research. But why should it be relevant to Philosophy of Mind then? Davidson says "the test is designed to throw light on the nature of thought [...and it] can be applied to any object" ([1990b] 2004, p. 78). While Davidson interprets the test as a sufficient condition for thought, his statement also reveals another more important aspect that is echoed by Turing:

> "The whole thinking process is still rather mysterious to us, but I believe that the attempt to make a thinking machine will help us greatly in finding out how we think ourselves." (Turing [1951] 1999, p. 465)

So the Turing Test and the related task of constructing a machine which can pass it, is a practical device to learn more about the nature

9 – For historical reasons, the terminology here is easily confusing: The term 'Empiricism' ought to be treated very carefully as distinct from empirical. Davidson's theory is empirical, meaning observable evidence plays an important role in it, but it disagrees with the dogmas of Logical Empiricism (sometimes also called Logical Positivism), namely the analytic-synthetic distinction and reductionism (Quine 1951) and the dualism of scheme and content (Davidson [1973a] 2001). Davidson replaces the concepts of meaning and proximal stimuli by a holistic and empirical theory which treats language and knowledge as elementary intersubjective (Lepore and Ludwig 2009, p. 22–23).

of thought and communication. This resonates very much with Davidson's philosophy. For the essential paradigm shift in his philosophy about language is following Quine's rejection of Carnap's Logical Empiricism[9]. The objective of a theory of meaning for Davidson "is sought not in reductive analyses [of the meaning of expressions], but rather in showing how evidence can be marshalled in support of a theory[10] of interpretation for a speaker." (Lepore and Ludwig 2009, p. 9). Philosophy is not an a priori discipline but simply a more general empirical discipline. " 'Language is a social art' [...and] evidence for its acquisition and deployment must be intersubjective, and, hence, recoverable from overt behaviour" (Lepore and Ludwig 2009, p. 22–23). Therefore the search for a definition of meaning or of intelligence itself is rather uninteresting. A definition is only relevant as part of an empirical theory that can be judged by its successful application. (I'll talk more about this at the end of this section.) The search for a definition is therefore replaced by the search of an empirical theory of intelligence.

From this angle the task of programming a machine which can pass the Turing Test seems closely related to that of devising and testing an empirical theory of intelligence. At least if we accept that the essence of such a theory could be expressed recursively. Judging a computer as passing the test is not so different from judging a person as rational being after all. Moor argues:

> "I believe that another human being thinks because his ability to think is part of a theory I have to explain his actions. [...T]he evidence for the theory comes from the outward behavior of the person. [...T]here is no reason why knowledge of computer thinking can not arise in the same way. I can use the computer's behavior as evidence in assessing my theory about its information processing." (Moor 1976, p. 251)

Davidson and Turing agree that we can devise a scientific theory that describes the essential parts of our linguistic competence. In the case of Turing this is evident in his believe that there is a program that allows the computer to win the imitation game. For Davidson it is evident in the proposal of his Unified Theory which can capture the essence of linguistic competence and rationality. However there are some important distinctions between the nature of these approaches that will be highlighted in subsection 3.1.

LINGUISTIC ABILITIES AS THE ESSENCE OF INTELLIGENCE

What we have discussed so far depends on accepting Turing's first claim that communication abilities are representative of intellectual abilities. As mentioned in the introduction, Davidson agrees since he "sees our nature as linguistic beings as the key to the possibility of thought" (Lepore and Ludwig 2009, p. 1). This believe springs from his metaphysics" and from his epistemology. The later will be examined in more detail in subsection 3.2. However, Davidson emphasizes the interrogator's judgement as the essential point:

> "Turing was right, in my opinion, in taking as the only test for the presence of thought and meaning the interpretive powers and abilities of a human interpreter." (Davidson [1990b] 2004, p. 86)

As we will discuss in subsection 2.3, Davidson very intentionally changes the emphasis here, since he argues that the ability to be interpreted is not a sufficient, but a necessary condition for attributing thought — introspection into the working of the mind for example is *not* a sufficient condition for Davidson. For Davidson the essence of Turing's approach is that "instead of asking how the content of a concept [...] is thought of by the creature that has the concept [...], we ask [...] how an observer can size up the contents of the thoughts of another creature" (Davidson [2001] 2004, p. 137).

The third claim (Turing's Principle) is not of particular interest to Davidson, but because of his naturalistic ontology[12] he agrees: "A person is a physical object which [...] functions according to physical laws. So [...] there is no reason why an artificial object could not think[... . The real question is:] how much like us must an artifact be, and in what ways, to qualify as having thoughts?" ([1990a] 2004, p. 87).

The second claim (the appropriateness of the test) is what Davidson discusses and criticizes extensively in his essay 'Turing's Test' ([1990b] 2004). The fourth claim (the importance of Machine Learning) is something Davidson does not pay much attention to, but that is closely related to his critique. Both are the topic of the next subsection.

So we have seen why *Davidson* finds Turing's Test particularly interesting. *I* have chosen his critique for discussion here since it clearly relates to Computer Science and because it "opens the way for Davidson's own view into the nature of thought" (Cavell 2004, p. xvii).

2.2 Davidson's Critique

Davidson finds Turing's Test inadequate to show that an object is thinking. Not because communication is not a sufficient criterion for intellectual competence for Davidson — as we have discussed. Neither because a test would require an introspection into the workings of the mind. Davidson explicitly says that it is not inadequate because it "restricts the available evidence to what can be observed from the outside" ([1990b] 2004, p. 83). But because it does not enable the interrogator to observe a history of three-way engagement between the object, a shared world, and other minds in which the object develops its semantics.

This is related to the "fundamental difference between semantics, which relates words to the world, and syntax, which does not" (Davidson [1990a] 2004, p. 94, he adopts the terminology from Tarski). For Davidson, the essential requirement for thought or intelligence is to assign meaning to words, to relate them to the world. But he argues

---

10 – Davidson moved from his interest in the historic aspect of philosophy to analytical philosophy because Quine had convinced him "that it was possible to be serious about getting things right in philosophy" and to take philosophy "as serious as science" (Davidson and Lepore [1999] 2004, p. 239). But through the recession of Logical Empiricism it had become clear that this progress could not simply lie in treating philosophy as the logic of mathematics and reasoning. His experiences in studying business, working with J.C.C. McKinsey on decision theory, and studying Tarski, gave Davidson "an appreciation for what it's like to have a serious theory" (Davidson and Lepore [1999] 2004, p. 253) and convinced him that the progress of ordinary language philosophy could be found in looking for a pragmatic empirical theory.

11 – Davidson argues that "If we have the semantics of a language right, the objects we assign to the expressions of the language must exist" (Davidson [1993] 2005, p. 40). Since being able to communicate successfully means getting semantics pretty right. This will in turn also mean that if we can communicate successfully, we must get it pretty right what things are in the world and in which relation they stand to each other. And this certainly seems to be a sufficient condition for intelligence. Davidson's argument for his claim that semantics is a method for metaphysics ([1993] 2005) is based on the fact that we must get things mostly right about the world, if we are intelligible as rational beings at all.

12 – Davidson's ontology takes only objects and events as basic entities ([1991b] 2001). He rejects the need for any propositions of proximal stimuli as entities in the mind ([1990] 2005). The content of propositional knowledge is directly caused by the distal objects and events the world is made up of.

that the interrogator in the Turing Test can not guarantee that the computer is able to do that:

> "[T]he interrogator [...] has no clue to the semantics of the object. There is no way he can determine the connection between the words that appear on the object's screen and events and things in the world. Of course there must be some connection; there is no other way to account for the intelligibility of the object's English. [...But i]t is perfectly possible that the connection between words and things was established by someone who programmed the object, and then provided purely syntactic connections between words for the object to wield. In this case it is the programmer who [...] has given meaning to the words [...], but the object doesn't mean anything, and there is no reason to take it to be thinking.
>
> In order to discover whether the object has any semantics, the interrogator must learn more about the connections between the output of the object and the world [...], through observing relevant causal interactions between the speaker, the world, and the speaker's audience. [...T]he interrogator [must be allowed] to watch the object interact with the world." (Davidson [1990b] 2004, p. 83)

The interpreter can only determine whether the computer means something by its words if he or she can tell what the computer means by them. Therefore "any evidence that thinking is going on will have to be evidence that particular thoughts are present" ([1990b] 2004, p. 80). While it is clear that thoughts have caused the computer's interaction, the test framework is not giving the interrogator a chance to determine how any thought has come to the computers knowledge, only that it has.

Davidson accredits this failure to the fact that "Turing wanted his test to draw '... a fairly sharp line between the physical and the intellectual capacities of man' (p. 434). [But t]here is no such line" (Davidson [1990b] 2004, p. 84). Turing was wrong that the 'body' of the object does not matter. Even though the details in which the sense organs convey impressions may not matter, the existence of such organs can not be reduced to purely textual communication. Turing has a somewhat Cartesian[13] approach to thought. He might not prescribe to an ontological mind-body dualism, but he tends to view thoughts as independent from sensory access to the world and only dependent on being able to communicate. Turing claims the "example of [the deaf blind] Helen Keller shows that education can take place provided that communication in both directions between teacher and pupil can take place" (Turing 1950, p. 456). He overlooks that Helen Keller for one had been able to see and hear for her first 19 months and, more importantly, only learned to communicate when her teacher correlated her signs to Helen's feeling of touch (Wikipedia 2015). So this example might serve more to illustrate Davidson's externalist approach to epistemology rather than underline Turing's point that communication alone is essential for thought.

The ability to interact with the world in any way and a history of such interactions and communications about those experiences with others is essential for developing a semantics. You "don't understand a language if there are not numerous connections between your use of words and experiences" (Davidson [1990b] 2004, p. 85). The interrogator needs to observe the computer's history of engagement with the world to judge its intelligence.

FAIRNESS OF THE CRITIQUE

So Davidson's main critique is that Turing's idea of a sharp line between physical and intellectual abilities is ill-conceived and Turing doesn't see the importance of interaction in a shared world as essential for (1) *developing* thoughts and for (2) *judging* intellectual abilities.

While the second critique is certainly justified, since the computer is hidden from the interrogator, the first point is debatable in light of Turing's work about Machine Learning. In his last footnote, Davidson claims that Turing "views this [learning] simply as an economical way of producing a device with mature thoughts; he does not see it as the only way" ([1990b] 2004, p. 86). As outlined in section 1, I think this is overestimating Turing's interest in presenting a sufficient condition for intelligence and underestimating his interest in actually building a certain grade of computer. Turing's ideas on Machine Learning take up a third of his essay (1950, p. 454–460) and constitute — as I argued — a fourth main claim. Most of his research from 1948–1952 was indeed not focused on the test which became so famous, but on approaches to build learning computers and simulating neural networks. None of this work mentions any other approaches for building AI. Turing actually outlines ideas how a computer with an indexed memory can be constructed that relates past experiences to new situations using associative connections and evaluation based on reactions of its teacher and later based on self constructed norms (see Turing [1951] 1996, p. 257–258). This seems closer to Davidson's requirement of having a history of interaction in a shared world than he gives credit for. So maybe classifying Turing as having a Cartesian approach to epistemology is premature.

Nevertheless, Davidson's critique of the *test* is justified. In the light of our discussion of Turing's work, however, our perspective might change, from seeing the inadequacy of the test as a result of Turing's underestimation of the importance of learning and interaction with the world, to surprise that he deemed the learning process, that he saw as essential too, as so irrelevant for the judgement.

2.3 Davidson's Proposal for a Modified Test

Davidson in fact, does not simply critique the test, but he proposes a modified test:

> "The Test must be modified [...]. The object must be brought into the open so that its causal connections with the rest of the world as well as with the interrogator can be observed by the interrogator.
>
> Can the interrogator now tell what the object thinks? The answer is that it depends [...]. Let us suppose the interrogator finds that the object uses words just as he does [...and] infers (let us suppose correctly) that the object's linguistic dispositions are similar to his own in relevant ways. In the case of a person, the interrogator would be justified in assuming that these dispositions were acquired in the usual way: in the basic cases, by past causal intercourse with things and circumstances of the sort to which the person is now disposed to respond. [...] But the assumption is not justified in the case of a computer: [...] The computer which has never experienced a dog and has no memory of dogs can't mean dog by the word 'dog'

13 – In fact Descartes somewhat preconceived Turing's test in his 'Discourse on the Method'. He claims, machines "could never use words or other signs arranged in such a manner as is competent to us in order to declare our thoughts to others" (Descartes [1637] 1993, Part V).

[...]. Thought and meaning require a history of a particular sort. [...U]nless we [...] can observe it in action over time, we have no basis for guessing how a computer came to have the dispositions it has.

It is unclear exactly what kind of history is necessary [...]. But our intuitions are clear enough in many cases. You [...] don't understand a language if there are not numerous connections between your use of words and experiences[. ...] It may seem that minds are, after all, inscrutable if no present observation of their operation can reveal what they are thinking. But of course this does not follow. [...E]ven the mind of an artefact can, if it has one, be understood; it just takes longer, long enough for some history to be observed, since it cannot be inferred." (Davidson [1990b] 2004, p. 84–86)

Davidson's rejection of the "sharp line between the physical and the intellectual capacities" ([1990b] 2004, p. 84) should not be misinterpreted as a rejection of the possibility to devise a test at all. In his follow-up essay 'Representation and Interpretation' ([1990a] 2004) he comes back to the question "what could we detach from a person and still count him or her as a thinking creature" (p. 87). And agrees with Turing that origin, building material, and size and shape for example are irrelevant. In so far, Turing was on the right track with his test, he only went too far in detaching the history from a person. The ability to determine non-predefined meaning — to connect symbols with objects and events in the world — is essential for intelligence and it depends on a rich base of experiences of causal interactions with these events and objects.

I put Davidson's explanation of the required nature of the test as follows: We may judge an object as intelligent, if (and only if — as I will argue) the object can be observed to be able to

1. successfully communicate with other intelligent beings and
2. derive its own semantics from a history of its experiences of interaction with other intelligent beings and with objects and events in a shared world.

This is what I call 'Davidson's Test'.

The ability to come up with its own semantics from a history of interaction ensures that the computer has a rich conceptual system in Davidson's holistic approach. Because "to have even one thought — one belief or desire — a computer would have to have a very great many other thoughts and desires. Beliefs and desires can exist only in the context of a very rich conceptual system."(Davidson [1990a] 2004, p. 90)[14].

We will see in how far this test is an expression of Davidson's theory of linguistic and intellectual competence throughout the next section and specifically in subsection 3.4.

INTERPRETATION OF DAVIDSON'S TEST AS A DEFINITION

So we have a definition of Davidson's Test, it remains to explain how it is to be interpreted. Davidson takes the Turing Test as aiming "to discover whether a sufficient condition for thought is satisfied; the condition is not claimed to be necessary" (Davidson [1990b] 2004, p. 81). So we can certainly interpret Davidson's Test as a sufficient condition for intelligence. But Davidson seems to go even further:

> "[T]he *only* way to tell if an artificial device [...] has [thoughts...] and the ability to perceive and interact with the world as a person does, is to attempt to *interpret the behavior* of the device in the same way we do the behavior of a person. [...U]nderstanding the program and physics of a device [...on the other hand] is *not* [...sufficient for] understanding the thought [...] of that device." (Davidson [1990a] 2004, p. 99, emphasis added)

It seems to me that this is interpreted correctly as the claim that Davidson's Test provides not only a sufficient, but also a necessary condition — and therefore constitutes an operational definition of intelligence. Interpreting the behavior of a person is nothing else for Davidson than interpreting its utterances (or some other communications with semantic content). This follows from the fact that the interpretation of the semantic content of an expression requires the interpreter to have a theory about the believes and semantics of the person and this theory must be based on the behavioral evidence the interpreter has. We will investigate this theory of interpretation from Davidson in more detail in subsection 3.3.

The second part of Davidson's claim aims to reflect an obvious counter-argument that might occur. If one completely understood the inner workings of a device's or person's brain, one could certainly also judge wether it was intelligent. While Davidson does not want to deny that dissecting a person's brain might enable one to undoubtedly judge a person as belonging to the species homo sapiens and therefore be justified to infer that he or she is intelligent, this is not the kind of proof we are looking for. The test provides an *explicit* definition of intelligence, while there might be other *implicit* ways of inference about intelligence. Davidson *does* reject that any knowledge of the brain can *explicitly* proof intelligence. To illustrate this he takes the programming of a machine as an example, which I will discuss in the next subsection.

I have argued before that the search for a definition is replaced by the search for an empirical theory in Davidson's philosophy. So if we talk about an operational definition here, this only becomes relevant holistically because it is part of an empirical theory of intelligence — and in Davidson's case this means a theory of interpretation.[15] So this theory is expected to fit empirical data about interpretations, to allow us to apply it and actually yield correct interpretations of utterances, or (since the theory is not developed far enough yet for that) at least to fit our intuitions for how we interpret others. Therefore we can't

---

14 – Davidson gives the following example to illustrate his point: "Alitalia Flight 19 leaves Turin for London on Tuesdays at 8:30 in the morning. We can learn this by consulting a computer; but does the computer know what we learn by consulting it? The answer is that it does not because it does not know what a flight is, where Turin is, or even that Tuesday is a day of the week." ([1990a] 2004, p. 89)

15 – To illustrate this point, let me give an example from first semester physics: Take momentum in classical physics — the fact that we single out the meaning of this word by attributing it as a part of the empirical theory of classical mechanics already shows that it gains its relevance from its place in this theory. It gains its relevance from its place in this empirical theory that serves to describe the movement of mass under natural and artificially applied forces. We can't debate the definition of momentum ($\vec{p}$), not because it is defined in very clear mathematical terms as the time derivative of the product of mass and location ($\vec{p} - \partial_t (m \cdot \vec{r})$), but because changing its definition would destroy some of the predictive power of classical mechanics. If we would for example only take the simplified definition of momentum as the product of mass and velocity (which assumes the mass doesn't change over time ($\partial_t m - 0 \Rightarrow \vec{p} - m \cdot \partial_t \vec{r} - m \cdot \vec{v}$)) — which is still very clearly defined mathematically — we couldn't explain how a rocket flies that accelerates by throwing away mass ($\partial_t m \neq 0$); the theory of classical mechanics would lose some of its power to fit empirical evidence. This is closely related to Popper's (1935) requirement of falsifiability for a scientific theory. The definitions are only good, as long as changing them would mean that the theory as a whole can't fit some empirical evidence anymore. In this sense their meaning and relevance is determined holistically.

judge this definition by itself; we need to look at Davidson's theory as a whole and see whether it is a convincing empirical theory for our linguistic and intellectual competence; then we can judge whether this definition of intelligence, which provides the goal for this theory, is convincing to us. Hence we will discuss Davidson's theory in more detail in the next section.

SUMMARY

We have started with a discussion of how Davidson's approach to language as a social art and interpretation as an empirical process relates to Turing's approach to intelligence, pointing out that both Davidson and Turing think that intelligence can be described empirically and that there is a relation between the process of building an intelligent machine and devising an empirical theory of intelligence. I have also argued that both see linguistic abilities as essential for intelligence and agree in their affirmative answer to the question wether artificial intelligence is possible, even though both do not find the question very interesting.

Than we have learned that Davidson critiques Turing's Test for not exposing how the computer develops its own semantics from a history of experiences of interactions in a shared world. I have argued that this shortcoming of Turing's design is surprising because he agrees with Davidson that learning from experience is essential for developing intelligence.

At the end I have given a characterization of Davidson's Test that requires not only the observation (1) of successful communication but also (2) of the process of deriving semantics from a history of interactions in a shared world. I argued that this test should be interpreted as an operational definition of intelligence which we can only judge in light of how empirically convincing Davidson's theory of linguistic and intellectual competence is for which it provides the goal.

3 Davidson's Theory of Intellectual and Linguistic Competence

The goal of this section is to ground all the arguments from before in an understanding of the main characteristics of Davidson's view of (3.1) the relation between the mental and the physical, (3.2) his epistemology, and (3.3) his theory of interpretation. After this I am going to discuss how those understandings come together in his unified theory of intellectual and linguistic competence and relate to Davidson's Test. Wrapping up with a discussion how Davidson's view might fit into the more recent discussion between connectionism and the classical representational and computational theory of mind.

3.1 The Meaning of Programs and Anomalous Monism

We will not tackle Davidson's view of the relation between the mental and the physical heads on, but follow his approach in 'Representation and Interpretation' ([1990a] 2004). Davidson starts out with the question "how much like us [a computer] must [...] be, and in what ways, to qualify as having thoughts" ([1990a] 2004, p. 88). But after establishing that a computer "cannot have thoughts unless it can learn and has learned from causal interactions with the world" ([1990a] 2004, p. 88) and that "what must be added to computers [...] to insure that they are capable of thought" ([1990a] 2004, p. 89) — that they know *what* they are talking about — is "a very rich conceptual system" ([1990a] 2004, p. 90). This is of course related to the holistic approach to semantics that we have discussed in subsection 2.3. Davidson's main question for the rest of the essay now becomes *how* we can judge that a computer has such a rich conceptual system. Davidson starts by ruling out that knowledge of the operation of the system is sufficient for this judgement:

> "It may seem obvious that if an artificial object thinks and acts enough like a person, someone who knows how the object was designed and built would be able to describe and explain the mental states and actions of the object. But this does not follow, for there is no reason to suppose that there are definitional or nomological connections between the concepts used by the designer and the psychological concepts to be described and explained. This should be clear if we imagine that the builder has simply copied, molecule by molecule, some real person. [...] So one sort of 'complete' understanding does not necessarily imply another." (Davidson [1990a] 2004, p. 90)

The problem is one of *mental representation*. The "representation of an object or a fact [...in] the program cannot automatically be interpreted as a representation of that object or fact for the device" ([1990a] 2004, p. 91) — in its 'mind'. Davidson argues that indeed the representations differ. If an object or event falls under a certain mental concept there must not be an equivalent concept in the program or in a physical description. He gives the following analogy: "Suppose, following folk advice, I am attempting to go to sleep by counting sheep. Every now and then, at random, a goat slips into the file. In my drowsy state I find I cannot remember the classificatory words 'sheep' and 'goat'. Nevertheless I have no trouble identifying each animal: there is animal number one, animal number two, and so on. In my necessarily finite list, I can specify the class of sheep and the class of goats: the sheep are animals 1, 2, 4, 5, 6, 7, 8, and 12; the goats are animals 3, 9, 10, and 11. But these classifications are no help if I want to frame interesting laws or hypotheses that go beyond the observed cases, for example, that goats have horns. I can pick out any particular sheep or goat in my animal numbering system, but I cannot, through conceptual poverty, tell the sheep from the goats generally. So it may be with the mental and the physical. Each mental event, taken singly, may have (must have, if I am right) a physical description, but the mental classifications may elude the physical vocabularies." ([1990a] 2004, p. 92)

The program is purely syntactic, it can give explanations of its concepts in terms of other concepts it has, but it cannot specify anything about semantics, about references of its concepts, about relations of its concepts to objects or events in the world. "There is the language in which each animal can be picked out, but which lacks the concepts needed for classifying the animals as sheep or goats; similarly, syntax can provide a unique description of each true sentence, [...] but it can't classify sentences as true or false. [...] If we knew no more than the program, we would have no reason to say [...] that any aspect of or event in the device represented anything outside the device" ([1990a] 2004, p. 91). Physics work in the same way, when it appeals to strict laws which are only "drawing on concepts from the same conceptual domain and upon which there is no improving in point of precision and comprehensiveness. [...] Physical theory promises to provide a comprehensive closed system guaranteed to yield a standardized, unique description of every physical event couched in a vocabulary amenable to law" ([1970] 2001, p. 223–224). Just like the program a complete physics is precise and contains no ceteris